DOCUMENT RESUME

ED 223 677 TM 820 806

AUTHOR Powers, Donald E.
TITLE Estimating the Effects of Various Methods of Preparing for the SAT.
INSTITUTION College Entrance Examination Board, New York, N.Y.; Educational Testing Service, Princeton, N.J.
REPORT NO CE-R-82-2; ETS-RR-82-23
PUB DATE 82
NOTE 29p.
PUB TYPE Reports - Research/Technical (143)

EDRS PRICE MF01 Plus Postage. PC Not Available from EDRS.
DESCRIPTORS *College Bound Students; *College Entrance Examinations; High Schools; Home Study; Independent Reading; Motivation; Review (Reexamination); *Test Coaching; *Test Results; *Test Wiseness
IDENTIFIERS *Scholastic Aptitude Test; Test Repeaters

ABSTRACT

A national sample of Scholastic Aptitude Test (SAT) candidates was surveyed to determine the ways in which they had prepared for the examination. Candidates' SAT scores were retrieved from test files, along with selected background information that included rank in high school class and scores on the Preliminary Scholastic Aptitude Test (PSAT), which most students had taken about 8 months before they took the SAT. Exploratory analyses of the effects of combinations of methods of preparation showed that various methods may interact in complex ways with other methods and with some student traits. Furthermore, self-selection may not be uniform across levels of ability and previous achievement and may depend on the candidates' selection of other methods of preparation. It is suggested that independent study (of either English or mathematics, or of test preparation books and test familiarization materials) may have an impact on test scores that is possibly as great as the effect of less-widely-available modes of preparation. It is also suggested that motivation may be a potentially important self-selection factor in quasi-experimental studies of special preparation. (Author/PN)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED223677

College Board Report

No. 82-2

Estimating the Effects of Various Methods of Preparing for the SAT

Donald E. Powers

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

X This document has been reproduced as received from the person or organization originating it.

Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

"PERMISSION TO REPRODUCE THIS MATERIAL IN MICROFICHE ONLY HAS BEEN GRANTED BY

S. W. Gardner

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC) AND USERS OF THE ERIC SYSTEM."

Estimating the Effects of Various Methods of Preparing for the SAT

Donald E. Powers

Educational Testing Service

College Board Report No. 82-2

ETS RR No. 82-23

College Entrance Examination Board, New York, 1982

The author wishes to acknowledge the advice of Hunter M. Breland, Assistant Director for Program Research at ETS. The report benefited from thoughtful comments and reviews by Donald L. Alderman, Brent Bridgeman, Thomas F. Donlon, Rex Jackson, Samuel Messick, and Donald A. Rock. The author wishes to thank Richard H. Harrison for conducting the data analysis, Lorraine Simon for performing numerous project administrative tasks and preparing the final manuscript, and Miriam Godshalk for editing it.

Researchers are encouraged to express freely their professional judgment. Therefore, points of view or opinions stated in College Board Reports do not necessarily represent official College Board position or policy.

The College Board is a nonprofit membership organization that provides tests and other educational services for students, schools, and colleges. The membership is composed of more than 2,500 colleges, schools, school systems, and education associations. Representatives of the members serve on the Board of Trustees and advisory councils and committees that consider the programs of the College Board and participate in the determination of its policies and activities.

Additional copies of this report may be obtained from College Board Publications, Box 886, New York, New York 10101. The price is $4.

Copyright © 1982 by College Entrance Examination Board.
All rights reserved.
Printed in the United States of America.

CONTENTS

Abstract 1

Introduction 1

Methods 2

Data Collection 2
Data Analyses 2

Results 3

Description of the Sample 3
Analysis I 6
Analysis II 11
Comparison of Findings with Results of Previous Studies 17

Summary and Discussion 17

Random Assignment 18
Description of Preparation Methods 18
Assessing Factors in Self-Selection 19
Nontest Score Effects 19
Sample Selection 19
Concluding Remarks 19

References 20

Tables

1. Mean Scores of Several Groups of SAT Candidates on Selected Variables 4

2. Comparison of SAT Scores and Other Selected Variables for Students Using Various Methods of Test Preparation 4

3. Additional Characteristics of Examinees Who Used Various Methods of Preparation 5

4. Effects on SAT-V, SAT-M, and TSWE Scores Associated with Various Methods of Test Preparation 6

5. Summary of Significant Interactions 9

6. Patterns of Overlap of Use of Several Methods of Preparation 11

7. Regression Equations Used to Predict SAT-V, SAT-M, and TSWE Scores for Analyses of Residuals 12

8. Differences Between Actual and Predicted Scores for Examinees Who Used Various Methods of Test Preparation 12

9. Predictions of SAT-V Residuals for Each Method of Preparation from HSR, PSAT-V, PSAT-M, and the Use of Other Methods of Preparation 14

10. Predictions of SAT-M Residuals for Each Method of Preparation from HSR, PSAT-V, PSAT-M, and the Use of Other Methods of Preparation 15

11. Predictions of TSWE Residuals for Each Method of Preparation from HSR, PSAT-V, PSAT-M, and the Use of Other Methods of Preparation 16

ABSTRACT

A national sample of Scholastic Aptitude Test (SAT) candidates was surveyed to determine the ways in which they had prepared for the examination. Candidates' SAT scores were retrieved from test files, along with selected background information that included rank in high school class and scores on the Preliminary Scholastic Aptitude Test (PSAT), which most students had taken about eight months before they took the SAT.

Multiple regression analyses revealed effects of less than ten points on the 200-800 SAT score scale and less than two points on the 20-60 score scale for the Test of Standard Written English (TSWE) for each of the methods of test preparation. A significant number of negative effect estimates, coupled with significant differences between the background characteristics of preparing and nonpreparing groups, strongly suggested the influence of selection bias in the estimates. Exploratory analyses of the effects of combinations of methods of preparation showed that various methods may interact in complex ways with other methods and with some student traits. Furthermore, self-selection may not be uniform across levels of ability and previous achievement, and may also depend on the candidates' selection of other methods of preparation. The implications of the results for the design of further research on the effects of special preparation for tests are discussed.

INTRODUCTION

Considerable attention has been paid, and much of it recently, to factors related to scores on the Scholastic Aptitude Test (SAT). Much of this attention has been focused on the effects of various types of special preparation on SAT scores. Most of the studies have investigated programs offered in or by secondary schools (Alderman and Powers 1980; Coffman and Parry 1967; Dyer 1953; Evans and Pike 1973; Frankel 1960; French 1955; Marron 1965; Pallone 1961; Roberts and Oppenheim 1966), but some studies of the programs that are offered by proprietary schools or commercial companies have been conducted (Federal Trade Commission 1979; Whitla 1962). With the exception of one study of the effects of the independent use of a test familiarization booklet prepared by the test sponsor (Powers and Alderman 1979), there appears to have been little formal study of the effects of other preparation strategies, such as the independent review of subject matter in various areas, the use of commercially available books on preparing for the SAT, and taking of the SAT for practice.

Moreover, compared with the longstanding search for knowledge of how students' characteristics relate to their responses to various educational interventions (see, for example, Cronbach and Snow 1977), there has apparently been relatively little concern for the possible interactions between student traits and special test preparation as an educational treatment, even though there has been speculation that such variables as student motivation and previous test-taking experience might interact with various forms of special preparation.

Some studies have addressed the possibility of differential effects for different kinds of students. Dyer (1953), for example, in a study conducted at two secondary schools for boys, noted an interaction between coaching and concurrent enrollment in mathematics courses. He found that coaching had a greater effect on the SAT-mathematical scores of students who were not taking mathematics courses than on the scores of students who were enrolled in mathematics. French (1955) found a similar pattern of effects for male students, according to their concurrent exposure to mathematics courses, but he found a pattern in the opposite direction for female students; that is, coaching was more effective for females who were studying mathematics than for those who were not. A reanalysis of the data obtained for the Federal Trade Commission's 1979 study of commercial coaching schools (Messick 1980) uncovered a number of interactive effects on the SAT-verbal scores at one of the schools studied. Although no differential effects were noted for SAT-mathematical scores either for years of mathematics taken or for sex, the effect of coaching on SAT-verbal scores was significantly greater for a small group of black students than for nonblack students. The effect on verbal scores was also greater for students who reported low family incomes than for those students who reported high parental incomes.

Although students may prepare in various ways, and in more than one way (Powers 1981), few studies have considered the effects of alternative methods, much less the potential interactions between various methods of special preparation, that is, the possibility that the effects of one method might differ according to whether or not another method was also used. The interaction between coaching and concurrent enrollment in mathematics courses reported by Dyer (1953) and French (1955) provides a good example of an effect of special preparation that was apparently moderated by another concurrent activity. Thus, the extent of student involvement in other forms of special preparation would seem to be a potential moderator of the effects of particular special preparation activities. Of course, there is also the possibility that the effects of various kinds of preparation might be additive.

This study was undertaken to estimate the effects of alternative methods of test preparation on SAT scores. These estimates reflect both the effects of using each method of preparation and a component attributable to selection. That is, a number of personal characteristics (e.g., motivation) of students who chose particular ways of preparing for the examination may be reflected in their subsequent SAT performance, thus giving the *appearance* of effects resulting from the kinds of preparation in which these students engaged. Therefore, when the effects of various methods of preparation on SAT scores are dis-

cussed, we are dealing with the *combined* effects of the particular method of preparation *and* the effect of selecting that method. It was hoped that a comparison of the various methods of preparation would lead to a greater understanding of their effects and also of the bases on which candidates select various methods of preparation.

We also hoped to consider, by computing interaction effects, the possibility that special preparation of various kinds might have differential effects according to the students' levels of ability and previous achievement, or that effects might differ according to other concurrent student experiences, including, especially, exposure to other forms of special preparation.

Finally, and perhaps of greatest importance, this study was intended to highlight, by using real data, some of the methodological problems involved in the interpretation of results of studies of special preparation, especially those using quasi-experimental approaches. The anticipated better understanding of the factors that give the *appearance* of effects on test scores was thought to be useful in specifying the critical design requirements of future studies of special test preparation.

METHODS

Data Collection

As part of a study of the test familiarization booklet, *Taking the SAT*, a random sample of 2,024 high school juniors who registered for the June 1978 SAT administration was surveyed about their reactions to the new booklet (Powers and Alderman 1979). Among the questions asked was one about the various other ways in which students had prepared to take the SAT, including the following:

1. use of a very brief test familiarization booklet called *About the SAT* that was provided to all candidates by the test sponsor
2. receipt of the more extensive test familiarization booklet, *Taking the SAT*[1]
3. independent review of mathematics
4. independent review of English or vocabulary
5. independent review of material from courses other than English or mathematics
6. use of test preparation books, such as those that are available commercially
7. attendance at review sessions held in school
8. attendance at review or coaching sessions given outside of school
9. taking the SAT for practice.

The SAT scores of 1,860 examinees who actually took the examination in June 1978 were retrieved from score files and matched with the questionnaire responses of 1,375 of these students. Selected background data were also obtained from the Student Descriptive Questionnaire (SDQ), which a majority of SAT candidates complete when they register for the test. Scores on the Preliminary Scholastic Aptitude Test (PSAT), which many of the students in the sample took in October 1977, were retrieved and matched with other data. The present study is based on data from 825 candidates who had complete data for all measures.

Data Analyses

Multiple regression analysis was used to compute estimates of the effects of each method of preparation. The dependent variables were scores on the verbal and mathematical sections of the SAT and scores on the Test of Standard Written English (TSWE), which is administered with the SAT. The independent variables were the self-reported use or nonuse of each of seven methods of preparation, whether or not candidates had previously taken the SAT, and whether or not they had received the booklet, *Taking the SAT*. Two separate sets of analyses were conducted in which both the main effects and the interaction effects associated with each method of preparation were computed.

Analysis I

The first set of analyses involved simultaneously regressing each of three dependent variables on the set of dichotomous variables that represented the use of each method of test preparation. High school rank (HSR), PSAT-verbal, and PSAT-mathematical were included as covariates. The raw regression weight for each dichotomous preparation variable was interpreted as the effect of the variable.

The inclusion of variables reflecting the students' use of alternative methods of preparation was thought to be important for two reasons: (1) their inclusion permitted an estimate for each method of preparation, and (2) the estimate of the effect of using a particular method of preparation might be less biased if adjustments were made according to whether or not other methods of preparation, which might reflect similar self-selection biases, were also used. That is, the candidates' use of various alternative methods of preparation might serve as proxies for a host of factors underlying the selection of each other method. Their inclusion as covariates might, therefore, help reduce the bias resulting from self-selection.

Next, interactions were assessed as follows. Each of the three dependent variables was, in turn, regressed on the set of all independent variables: PSAT-verbal scores, PSAT-mathematical scores, HSR, and all of the nine variables reflecting the students' involvement in various kinds of preparation. Product terms were formed to represent the various interactions of interest, which included those between student traits (e.g., verbal aptitude as measured by the PSAT-verbal section) and the students' methods of preparation (e.g., independent review of English), and

1. For the Powers and Alderman (1979) study of *Taking the SAT*, half the 2,024 candidates in the sample were mailed copies of *Taking the SAT* about four weeks before the examination; the other half were sent no special materials.

also those between each pair of methods of preparation (e.g., independent review of mathematics and coaching outside of school). Each interaction term was in turn added separately to the regression system containing PSAT-verbal scores, PSAT-mathematical scores, HSR, and the preparation variables, and the significance of the regression weight for each product variable term was noted. For those pairs of variables that interacted significantly ($p < .01$), the effect of each variable was computed at each of two values of the other variable. For the dichotomously coded variables indicating involvement in each method of preparation, the interaction effect was estimated by comparing the predicted criterion scores of the four groups of examinees according to their use or nonuse of each of the two methods of preparation of interest, while assuming constant values for all four groups on the other covariates.

Similarly, to assess the interactions between the dichotomous preparation variables and the nondichotomous variables such as the PSAT, the effect of each method of preparation was computed at two levels of the other variable, at one standard deviation above the mean of all examinees, and at one standard deviation below, while a constant value for all other variables was again assumed.

Finally, those interactions that were found to be significant when entered alone with the other covariates were then entered simultaneously with the covariates in order to check the extent to which these interaction effects might account for the same variation in test scores, thereby receiving less weight in the regression equations. The results discussed below are based on the separate, not simultaneous, entering of interaction terms because of the slightly less cumbersome computations required and because of the difficulty in interpretation when a particular variable interacted with several others.

Analysis II

Because of the nature of the overlap of the candidates' use of various methods of preparation, a second set of analyses was undertaken to assess the robustness of the results of Analysis I. By overlap is meant that most candidates used several methods of preparation and very few failed to prepare at all, thus rendering a no-preparation control group impractical. For this second set of analyses, a group of students ($N = 168$) who had used only the materials provided by the College Board was identified as a baseline group against which the effects of each of the other test preparation methods could be gauged. The rationale for choosing this control group was that, whereas some methods of test preparation—commercial coaching in particular—are not accessible to all students, the materials provided by the College Board are freely available to all students.

Three separate regression equations were computed for this baseline group for each of the three dependent variables, again using high school rank and PSAT scores as predictors. The first prediction was based only on high school rank and PSAT scores. The second included these variables and also variables indicating which of the College Board materials these students had used. The third equation allowed for a nonlinear prediction by permitting quadratic terms for high school rank and PSAT scores.

Next, for students who used each of the preparation methods, differences between actual SAT and TSWE scores and those predicted from each of the three baseline-group regression equations were obtained. For each preparation method, the average of these residual scores was interpreted as the effect of using and/or selecting to use the particular method.

Interaction effects were also computed by regressing the residuals for each method, that is, the individual student effects, on other variables, including the use of other methods of preparation. Significant correlations of these individual effect estimates with background variables or with the use of other test preparation methods could be interpreted as an interaction effect (Stroud 1980).

Each set of analyses was restricted to cases for which complete data were available ($N = 825$). The decision to use a limited number of covariates (PSAT scores and high school rank) with the variables indicating the use of various preparation methods was reached by considering the number of complete data cases lost by adding variables. A more important consideration was that other previous studies had suggested that three covarites (PSAT-verbal scores, PSAT-mathematical scores, and HSR) would account for a very high proportion of the variation in SAT scores. Some variables for which data were available, such as self-reported latest grade in English or mathematics, were excluded not only because this information was reported less frequently, thus resulting in more missing data, but also because some kinds of preparation (e.g., review of mathematics) might have affected these variables, making them less than strictly appropriate as covariates.

Although all information from the SDQ was not used in the multiple regression analyses, chi-square tests were used to assess the significance of the differences between the SDQ responses of those who used and those who did not use each method of preparation. This analysis was intended to provide further clues as to the nature of the choice of special preparation methods. Preparers and nonpreparers were also compared with respect to their responses to the questionnaire used in the study of *Taking the SAT*.

RESULTS

Description of the Sample

The results presented here are based on a sample of 825 June 1978 SAT candidates for whom complete data were available with respect to PSAT-verbal and PSAT-mathematical scores, self-reported rank in high school, questionnaire responses pertaining to their use of various alternative methods of test preparation, and SAT and TSWE scores. To assess the degree to which this final study sample may be

atypical of SAT candidates in general (and the results therefore limited in their generalizability), see Table 1. This table shows mean scores for PSAT, SAT, and high school rank, where available, for all SAT candidates in 1977-78, and also for the sample on which the results reported here are based. As can be seen, the general effect of selection at each of the various stages was to render a final sample of candidates who were more able, in terms of their test scores and rank in high school, than candidates in general.

Table 2 shows the number and percentage of students who used each method of preparing for the SAT (and the number who did not use any methods), their high school ranks, and their scores on the PSAT, the SAT, and the TSWE. Clearly, there is substantial variation in the extent to which examinees used each of the methods to prepare for the test. That many examinees necessarily used more than one method is clear also.

Differences among each preparation group were ap-

Table 1. Mean Scores of Several Groups of SAT Candidates on Selected Variables

Variable	*All SAT Candidates (N = 1,417,232)*[a]	*All June 1978 SAT Candidates (N = 176,814)*[a]	*Taking the SAT Study Sample (N = 1,860)*	*Taking the SAT Study Questionnaire Respondents (N = 1,334)*	*Taking the SAT Study Respondents with Available PSAT Scores and HSR (N = 825)*
SAT-V	429	428	436	449	461
SAT-M	469	481	493	507	521
TSWE	42.8	43.2	44.5	45.7	46.8
PSAT-V	n.a.	n.a.	n.a.	n.a.	44.2
PSAT-M	n.a.	n.a.	n.a.	n.a.	48.5
HSR	76.0	n.a.	76.1	77.9	79.7

Note: n.a. = not available.

a. Source: College Board Admission Testing Program Statistical Summary for Academic Year 1977-78.

Table 2. Comparision of SAT Scores and Other Selected Variables for Students Using Various Methods of Test Preparation

				Independent Variables			*Dependent Variables*		
Method of Preparation	*N*	*Percent*[a]		*PSAT-V*	*PSAT-M*	*HSR*[b]	*SAT-V*	*SAT-M*	*TSWE*
Completing the questions in *About the SAT*	649	78.7	M	43.9	48.2	79.8	459.4	520.5	46.9
			SD	9.9	9.7	15.9	104.6	109.0	9.7
Reviewing mathematics independently	223	27.0	M	43.3	46.4	78.5	447.4	501.2	45.3
			SD	11.0	9.4	16.2	108.7	108.0	10.4
Reviewing English or vocabulary independently	354	42.9	M	43.8	48.3	79.0	460.9	518.9	46.7
			SD	10.6	9.9	16.2	109.2	113.5	10.1
Reviewing material from other courses independently	62	7.5	M	41.3	46.1	76.7	433.1	489.8	44.2
			SD	11.0	10.8	16.5	109.2	132.7	10.8
Using test preparation books	423	51.3	M	43.6	47.9	79.2	460.0	514.1	46.3
			SD	10.3	9.6	16.2	109.2	109.5	10.2
Attending review sessions in school	138	16.7	M	42.7	46.6	76.5	437.0	498.4	44.9
			SD	9.3	9.5	16.1	90.3	101.8	9.3
Attending review or coaching sessions outside school	27	3.3	M	42.4	44.7	75.7	440.7	481.9	41.9
			SD	9.3	10.0	18.5	95.9	115.9	10.6
Taking a previous SAT	68	8.2	M	44.0	48.2	82.4	458.5	521.5	46.8
			SD	8.8	9.0	13.5	97.3	107.1	9.0
Receiving the booklet, *Taking the SAT*	380	46.1	M	44.4	49.0	79.9	465.0	529.9	47.4
			SD	9.9	9.4	16.0	104.7	104.2	9.7
No preparation	41	5.0	M	45.3	50.6	80.9	462.2	543.2	46.8
			SD	11.1	10.2	17.7	104.3	105.2	8.3
All students	825	100.0	M	44.2	48.5	79.7	461.3	521.5	46.8
			SD	10.0	9.7	16.2	104.8	107.4	9.7

a. In terms of 825 students having complete data, i.e., SAT scores, PSAT scores, HSR, and questionnaire responses.

b. Response categories were class standing in fifths and tenths. These responses have been converted to percentile ranks using the midpoint of each category.

parent with respect to group standing on pre-SAT variables. When compared with all 825 students in the sample, the students who reported using each particular method of preparation exhibited a tendency for lower standing (in most cases only slightly) in terms of PSAT scores and high school rank. Students who reported independently reviewing materials from courses other than English or mathematics had lower PSAT-verbal scores than any other group, as well as relatively low PSAT-mathematical scores and high school rank. The few students who attended coaching or review sessions outside of school had lower PSAT-mathematical scores and rank in class than any other group. They also had relatively low PSAT-verbal scores. Although the students who had previously taken the SAT exhibited scores at almost exactly the average of the total sample, their overall average standing in high school was higher than that of any other preparation group.

Table 3 describes other student characteristics (as reported on the SDQ) that were significantly related to the examinees' use of each method of preparation. Because of the number of comparisons, only those characteristics on which users and nonusers differed at the .01 level have been included. As can be seen, the most frequently noted difference between examinees who used a given method of preparation and those who did not was the degree of test-related nervousness reported at various times.

Examinees who chose to use *About the SAT*, to review vocabulary or English on their own, to review material from courses other than English or mathematics, or to use test preparation books, reported being more nervous than nonusers before, during, or after the test administration. With the exception of those who used *About the SAT*, who were significantly more nervous *during* the test, all other significant differences in nervousness occurred either before the test or while waiting for test scores. For three of the four preparation methods for which users were more nervous than nonusers, the differences in anxiety were to be restricted primarily to the period in which examinees waited for scores. Thus, this continued anxiety may reflect the importance of the test to these examinees, and therefore may be different in quality from the test anxiety to which impaired test performance is sometimes attributed.

Other significant differences included the following: Examinees who reviewed either mathematics or English and vocabulary independently were more likely to be members of a minority group than were those who did not review. This finding holds true for minority students as a single group, and for each of five specific minority groups. Those who attended coaching or review sessions outside of school appear to have been selected, or to have selected themselves, at least partially on the basis of their mathematics history. These examinees reported taking fewer mathematics courses and obtaining lower grades in their most recently completed mathematics course. They also reported higher levels of education for their mothers. The only other characteristic on which preparers differed significantly from nonpreparers was self-reported mechanical ability; that is, those who reviewed material from courses other than English or mathematics reported themselves to be higher in mechanical ability than those who did not review such material.

In summary, the characteristics of the candidates choosing each method of preparation do indicate selection effects. That is to say, candidates choosing any given method of preparation are, generally speaking, different in certain self-reported characteristics from candidates who do not choose that particular method. This holds true for nearly every method studied here. Moreover, the characteristics on which preparers and nonpreparers differ are not necessarily the same for each method, which suggests that different bases may be used to select each method. Furthermore, it is not clear how each of these student characteristics is related to test scores, and therefore the direction or magnitude of any resulting selection bias also remains unclear.

Table 3. Additional Characteristics of Examinees Who Used Various Methods of Preparation

Method of Preparation	*Characteristics of Those Using the Method (Compared with Those Not Using the Method)*
Completing the questions in *About the SAT*	More likely[a] to report being more nervous during the test and while waiting for test scores.
Reviewing mathematics independently	More likely to be a member of a minority group.
Reviewing English or vocabulary independently	More likely to be a member of a minority group. More likely to report being more nervous when registering for the test, a few minutes before the test, and while waiting for scores.
Reviewing material from courses other than English or mathematics	More likely to report themselves high in mechanical ability. More nervous when registering.
Using test preparation books	More likely to report being nervous a week before the test, just before the test, and while waiting for scores.
Attending review sessions in school	—
Attending review or coaching sessions outside school	More likely to report lower most recent year-end grades in mathematics and fewer years of mathematics instruction. More likely to report higher level of education for mother.

a. All differences are significant at the $p < .01$ level using a χ^2 test.

Analysis I

Effects on SAT-Verbal, SAT-Mathematical, and TSWE Scores

The estimates of the effects of each method of preparation and the associated standard errors are shown in Table 4. We restate that each of these estimates actually represents the combined effect of the particular method of preparation and of deciding to use it. Examinees who choose (or are chosen) to use a particular method of preparation may be more likely than those who do not choose to use that method to obtain higher SAT and TSWE scores, even without using the particular preparation method, because they differ in some other ways that are related to test scores.

For each dependent variable—SAT-verbal, SAT-mathematical, and the TSWE—a large proportion of the variance was explained by high school rank, PSAT-verbal and PSAT-mathematical scores, and the nine preparation variables. For SAT-verbal and SAT-mathematical scores, the multiple correlations (.88 and .86) approach the limits allowed by the reliability of the measures. Such high correlations would suggest that, for SAT-verbal and SAT-mathematical scores, the amount of bias in estimates resulting from unmeasured correlates of selection may be relatively small. However, given the lower multiple correlation of the TSWE (.74) with the independent variables, the estimated effects on the TSWE are more likely to reflect selection bias. It should be noted that the corresponding multiple correlations obtained by using only PSAT-verbal and PSAT-mathematical scores and high school rank were as high, or nearly so (.88, .86, and .72), as those obtained by using all variables. In addition, the effects of each method of preparation were estimated with different degrees of precision, primarily because of the substantial differences among the sizes of the preparation groups. Standard errors of estimate were relatively small for the effects of receiving *Taking the SAT* (3.48 for SAT-verbal scores and 3.84 for SAT-mathematical scores), but they were larger for the effects of coaching outside of school (9.94 for SAT-verbal scores and 10.96 for SAT-mathematical scores).

Overall, none of the effects is very large. For SAT-verbal scores, in fact, the two largest estimates (for reviewing mathematics independently and for attending review sessions in school) are negative. Because it is difficult to believe that any preparation, such as attending review sessions in school, would result in lower SAT-verbal scores, it seems likely that these negative estimates reflect self-selection bias for these preparation methods. Negative estimates of this kind have been noted in a number of other contexts, particularly in evaluations of compensatory education programs (e.g., Campbell and Erlebacher 1970; Campbell and Boruch 1975), and such negative estimates have been attributed to the failure to adjust for other variables that are related to the covariates, to the dependent variables, and to group membership.

Table 4. Effects on SAT-V, SAT-M, and TSWE Scores Associated with Various Methods of Test Preparation

	Dependent Variable					
	SAT-V[a]		*SAT-M*[a]		*TSWE*[a]	
Method of Preparation	*Estimated Effect*	*Standard Error of Effect*	*Estimated Effect*	*Standard Error of Effect*	*Estimated Effect*	*Standard Error of Effect*
Completing the questions in *About the SAT*	3.32	4.28	9.78	4.72	1.44	0.57
Reviewing math independently	− 9.28	4.41	−1.29	4.87	−0.95	0.59
Reviewing English or vocabulary independently	7.04	4.00	2.26	4.41	0.75	0.54
Reviewing material from other courses independently	− 0.97	6.87	−7.97	7.57	−0.37	0.92
Using test preparation books	9.01	3.65	−3.11	4.02	−0.25	0.49
Attending review sessions in school	−11.94	4.70	−4.83	5.18	−0.63	0.63
Attending review or coaching sessions outside school	2.52	9.94	−2.75	10.96	−2.80	1.33
Taking a previous SAT	− 1.42	6.36	2.92	7.01	0.47	0.85
Receiving the booklet, *Taking the SAT*	2.78	3.48	8.26	3.84	0.70	0.47
Multiple Correlation	$R = .88$		$R = .86$		$R = .74$	

Note: HSR, PSAT-V, and PSAT-M were included in each regression equation with all of the preparation variables listed above.

a. SAT-V and SAT-M are on a 200–800 scale. TSWE is on a 20–60 scale.

The estimates in Table 4, therefore, whether they are positive or negative, may be either overestimates or underestimates of the effects of each method of preparation. However, with some consideration of the likelihood of various kinds of selection of methods, it may be possible at least to offer hypotheses about the direction, if not the magnitude, of the bias.

The largest positive effects on SAT-verbal scores were associated with the independent review of vocabulary or review of material from English courses (7.04 points) and with the use of test preparation books (9.01 points). Relatively large *deficits* were associated with the independent review of mathematics (-9.28 points) and with attending review sessions in school (-11.94 points). Again, it is important to note that each of the estimates reflects the effects of self-selection and of using the particular method of preparation. In contrast to the estimates given above, the data suggest that the effects on SAT-verbal scores are different for reviewing English courses and for using test preparation books than for either attending in-school review sessions or reviewing mathematics, a strategy that would not be expected to affect SAT-verbal scores.

The largest effects on SAT-mathematical scores were associated with completing the questions in the brief candidate booklet, *About the SAT* (9.78 points), and with receiving the more extensive familiarization booklet, *Taking the SAT*, which contains strategies for approaching questions and includes a full-length sample test (8.26 points). Since *Taking the SAT* was distributed to a random half of the total sample of students, the estimate for this method of preparation does not reflect self-selection. In fact, because the variable under consideration here is simply *receiving* the booklet, our analysis may underestimate the effect of actually reading the booklet to prepare for the test. Furthermore, since not all students who had read the booklet actually completed the full-length sample test it contains, the benefit of using the complete booklet, including the sample test, may be further underestimated. On the other hand, the estimate of the effect of using *About the SAT* may indeed reflect the effect of self-selection as well as the effect of use. It seems intuitively reasonable that the much less extensive *About the SAT* (9.78 points) would have less effect on scores than *Taking the SAT*. The nearly 8-point deficit associated with reviewing material from courses other than English or mathematics may reflect selection primarily, and a number of candidate characteristics that are related to test scores could be conjectured to support this hypothesis. The estimates of the effects on SAT-mathematical scores of each of the other methods of preparation are small.

With the exception of using *About the SAT* and attending review or coaching sessions outside of school, the estimates of the effects on TSWE scores of each method of preparation were less than one point on the 20-60 point TSWE scale. A positive estimate of nearly 1.5 points was computed for completing the questions in *About the SAT*, while a large deficit (-2.8 points) was calculated for attending review on coaching classes outside of school. This negative estimate may reflect the fact that the coached students report lower ranks in class than do students in any other group. They also (see Table 3) report receiving lower mathematics grades and taking fewer mathematics courses. These students, therefore, may be generally lower achieving than noncoached students. This apparent achievement-related selection could be expected to be reflected more by lower scores on the TSWE, an achievement measure, than by lower scores on the less achievement-dependent aptitude portions of the SAT.

Because *Taking the SAT* was made available randomly to half of a specified sample of SAT candidates, the estimates of the effects of receiving this booklet were *not* confounded with the effects of selection, and may, therefore, deserve some additional mention here. In general, the small, insignificant effect (2.78 points) on SAT-verbal scores and the somewhat larger, significant effect (8.26 points) on SAT-mathematical scores are consistent with previous studies showing a slightly greater effect of special preparation on SAT-mathematical than on SAT-verbal scores (Messick 1980). A small, insignificant effect (.70 point on the 20-60 point scale) was noted for the TSWE. The direction and magnitude of these estimates are more consistent with expectations based on previous research than are some of the other estimates.

If it is reasonable to assume that the effects of the self-selection of particular methods of preparation operate in a similar manner across SAT-verbal, SAT-mathematical, and TSWE scores, then we can compare the effect estimates for each method of preparation to determine the test section or sections for which the use of various methods is most likely to produce positive score changes. Needless to say, the reasonableness of this assumption can be challenged on several grounds. Self-selection related to motivation, or to any of several other factors, might be differentially related to test scores on the various sections of the SAT and the TSWE. For example, extraordinary motivation might be more likely to result in higher scores on the TSWE and the more curriculum-related SAT-mathematical sections than on the SAT-verbal sections, which may be less susceptible to change even when motivation runs high. And finally, selection based on a perceived or actual deficiency with respect to a particular ability, namely mathematical or verbal aptitude, would certainly operate differently for each section of the test. Nevertheless, examining the differential effects of the various preparation methods may provide some additional clues regarding the effects of self-selection and preparation.

The methods that appear to have had a greater effect on SAT-mathematical performance than on SAT-verbal performance were the reviewing of mathematics independently (a difference of 7.99 points) and attending preparation sessions in school (a difference of 7.11 points). Conversely, using test preparation books, reviewing English

course material, and reviewing material from courses other than English and mathematics seemed to have a differential effect favoring SAT-verbal scores. The fact that, as might be expected, reviewing verbal-oriented courses affects SAT-verbal scores more than it affects SAT-mathematical scores, and that reviewing mathematics courses affects SAT-mathematical scores more than it affects SAT-verbal scores, lends some credibility to these estimates and constitutes, possibly, some evidence of the efficacy of independent review. The largest of the differentials (12.12 points for using test preparation books) favoring SAT-verbal scores contrasts sharply with the differentials in favor of SAT-mathematical scores for the two familiarization booklets provided by the test sponsor. One possible explanation for this pattern is the difference in emphasis of, especially, *Taking the SAT* from the emphasis of many of the commercially available test preparation books. *Taking the SAT* contains a fairly extensive review of mathematics concepts and has relatively little emphasis on vocabulary development, whereas commercial test preparation books are much more likely to contain vocabulary lists.

There were also some large differentials between the effects on TSWE scores and the effects on SAT-verbal and SAT-mathematical scores. The effects were more pronounced, after considering differences in scales, for TSWE scores than for SAT-verbal scores for completing the questions in *About the SAT*, somewhat less for TSWE scores than for SAT-verbal scores for using test preparation books, and considerably less for attending coaching sessions outside of school. The effects on SAT-mathematical scores were greater than those on TSWE scores for reviewing mathematics and were also much greater for attending coaching. The deficit on TSWE scores is consistent with the hypothesis that coached students have been selected at least partly on the basis of low previous achievement, a selection factor for which an adjustment on the basis of self-reported high school rank and PSAT scores may not be adequate, and which therefore may appear as a deficit on the TSWE achievement test.

Several interpretations of interaction effects are possible. First, the effectiveness of a particular method of preparation may depend on the examinees' level on particular traits (trait-by-method interactions) or on their having participated in some other method of preparation (method-by-method interactions). As an example, high-ability examinees may be more likely to benefit from a particular method of preparation than lower ability examinees, or examinees who are deficient in previous test-taking experience may benefit more from a certain preparation method than more experienced test takers do, simply because the former students have more room for improvement.

A second major interpretation of interactions in the present study is that the effects of self-selection to particular methods of preparation may depend on the examinees' level on particular traits or on their having participated in other methods of preparation. For example, the reasons for, or correlates of, selecting a particular method of preparation may be different for candidates who score high on the PSAT than they are for those who score low. We could hypothesize that the former students might volunteer to attend review sessions conducted by their secondary schools because they are extremely motivated to obtain high SAT scores, whereas the latter students may have been actively encouraged by school officials to enroll in the review sessions in order to improve subsequent SAT performance, and they may, therefore, be more reluctant, less active participants.

A similar differential self-selection hypothesis can be advanced for method-by-method interactions. Examinees choosing a particular method of preparation might do so on a different basis, depending on whether or not they had also chosen another method of preparation. Those choosing each of two methods, in contrast to those choosing only one, may be, for instance, either more motivated or more anxious about test-taking, or they may simply have greater access to various test preparation resources and possibly to other educational opportunities as well.

Furthermore, since examinees may use methods of preparation serially rather than simultaneously, the examinees' selection of subsequent preparation may depend on their perceptions of the effectiveness of a preceding method. For example, examinees who perceive a particular method of preparation to be adequate may be more likely to forgo additional preparation than examinees who are less satisfied, or they may base their decisions to use other methods on some other factors.

In short, the issues of self-selection and of possible interactions between candidates' traits are complex. It is hoped that the exploratory analyses of interaction effects in the following section may produce some hypotheses concerning the apparent effects of selecting to use, and the actual effects of using, various methods of preparation. Table 5 presents all interactions that were significant at the .01 level.

Trait-by-Method Interactions

For Sat-verbal scores, two significant ($p < .01$) interactions between traits and methods of preparation were detected: PSAT-verbal scores with attending coaching outside of school and PSAT-mathematical scores with using test preparation books. For SAT-mathematical scores only one such interaction was significant: PSAT-mathematical scores with reviewing courses other than English or mathematics. No significant trait-by-method interactions were found for TSWE scores.

Interaction of PSAT-Verbal Scores with Coaching (for SAT-Verbal Sections). Although, overall, the effect on SAT-verbal scores of attending review or coaching sessions outside of school was negligible (about 2-3 points), the significant interaction effect suggest that either (1) that this method of preparation has differential benefits depending

Table 5. Summary of Significant Interactions

Method of Preparation	*Candidate Trait or Other Preparation Method*	*Dependent Variable*	*Description of Effect*
Coaching outside of school	PSAT-V	SAT-V	−37 points for high[a] PSAT-V candidates, +33 points for low[a] candidates.
Using test preparation books	PSAT-M	SAT-V	+18 points for high[b] PSAT-M candidates, −1 point for low[a] candidates.
Reviewing courses other than English or mathematics	PSAT-M	SAT-M	+13 points for high[b] PSAT-M candidates, −25 points for low[b] candidates.
Independent review of mathematics	Coaching outside of school	SAT-M	Review of math: −4 points if not coached, +62 points if coached. Coaching: −45 points if no math review, +21 points if math review.
Using *About the SAT*	Using test preparation books	SAT-M	*About the SAT*: −2 points if other books not used, +24 points if other books used. Test preparation books: −23 points if *About the SAT* not used, −2 points if *About the SAT* is used.
Independent review of English	Coaching outside of school	TSWE	Review of English: +0.6 points if not coached, +9.0 points if coached. Coaching: −8.6 points if no English review, −0.2 points if English review.
Using test preparation books	Attending in-school review sessions	TSWE	Test preparation books: +1.3 points if no in-school review, −3.1 points if in-school review. In-school review: +1.5 if no test preparation books, −1.9 if test prepration books.

Note: The reader should recall the exploratory nature of the analyses on which the tabled results are based. Because of the many uncontrolled factors, the results reflect both self-selection and actual effects of test preparation. Also, because of the larger number of unprotected significance tests that were made, the number of significant interactions is not very large, and therefore require replication before any conclusions can be drawn.

a. Low- and high-scoring PSAT-V candidates were defined as those at one standard deviation below (34) or above (54) the PSAT-V mean of all candidates in the study.

b. Low- and high-scoring PSAT-M candidates were defined as those at one standard deviation below (39) or above (58) the mean PSAT-M of all candidates in the study.

on the student's level of verbal ability as measured by the PSAT-verbal section, or (2) that the effects of self-selection vary with verbal ability or with PSAT-verbal scores, or with some combination of both. At one standard deviation below the average PSAT-verbal score of the total study sample, the estimate of coaching/review is about 33 points; at a level one standard deviation above the PSAT-verbal mean for the total group, the estimate is -37 points. Because it is difficult to think that coaching could negatively affect the SAT-verbal scores of verbally able examinees, self-selection may be the primary determinant of at least this negative estimate. Students may have chosen to attend coaching sessions because they obtained PSAT-verbal scores that were lower than expected, possibly because they were unlucky. This contention is supported by the fact that the coached students in this sample obtained lower PSAT-verbal scores than did any other group. Those candidates who received high PSAT-verbal scores, on the other hand, may not have suffered from bad luck, but in fact may have obtained PSAT-verbal scores that were somewhat higher than subsequent testings would suggest that they deserved.

Interaction of PSAT-Mathematical Scores with Test Preparation Books (for SAT-Verbal Sections). As noted above, a positive effect on SAT-verbal scores was associated with choosing to use test preparation books. A positive interaction effect suggests that this difference was concentrated in the high PSAT-mathematical score range. At one standard deviation below the total group PSAT-mathematical mean score, the effect was insignificant (-1 point), but at one standard deviation above the mean, it was about 18 points. It seems curious that those getting high PSAT-mathematical scores might benefit more by using test preparation books than their lower PSAT-mathematical-scoring counterparts might. A number of conjectures regarding this pattern are possible. For example, those getting high PSAT-mathematical scores may choose to concentrate on the verbal sections of test preparation books, whereas those getting lower PSAT-mathematical scores may choose to focus their study on the materials relevant to the mathematical sections of the test, at the expense of verbal preparation. This conjecture is supported somewhat by the fact that examinees scoring high on the PSAT-verbal section benefit somewhat more, though not significantly more, on the SAT-mathematical sections from using test preparation books than do low PSAT-verbal examinees. The reader may conjecture a number of other equally plausible and equally complex self-selection hypotheses.

Interaction of PSAT-Mathematical Scores with Reviewing Other Courses (for SAT-Verbal Sections). This interaction, too, seems curious—that a positive effect should be computed for students having high PSAT-mathematical scores, while a negative effect was associated with those having low PSAT-mathematical scores. At one standard deviation below and one standard deviation above the total sample PSAT-mathematical mean, the estimated effects were -25 points and +13 points. Deciding what specific kinds of course review other than mathematics might produce an interaction of this nature is difficult, although quantitatively oriented courses such as physics, chemistry, and mechanical drawing might sharpen certain skills (e.g., the ability to interpret data) or content familiarity (e.g., knowledge of geometric concepts) that are tested on the SAT-mathematical sections. Any score-related variables that would explain differential self-selection effects at different levels of PSAT-mathematical scores do not readily come to mind.

Method-by-Method Interactions

For SAT-verbal scores, no significant ($p < .01$) method-by-method interactions were detected, but for SAT-mathematical scores, three significant interactions were found: independent review of mathematics with coaching outside of school, independent review of English with coaching outside of school, and using *About the SAT* with using text preparation books. For the TSWE, there were two significant interactions: independent review of English with coaching outside of school and using test preparation books with attending in-school review sessions.

Interaction of Independent Review of Mathematics with Coaching (for SAT-Mathematical Sections). Negligible negative overall effects on SAT-mathematical scores were computed for independently reviewing mathematics and for attending review or coaching sessions outside of school. However, these two preparation variables interacted significantly ($p < .01$) in such a way that an insignificant negative effect (-4 points) of independent mathematics review was noted when there was no coaching, while a large (62-point) positive effect was observed for mathematics review when there was coaching. Alternatively, the effect of coaching was large and negative (-45 points) if coaching was not accompanied by independent mathematics review, but positive (21 points) if mathematics was reviewed also.

In the absence of any self-selection effects, one would conclude that independent mathematics review may help a good deal if students are also coached—possibly, we might guess, because the coaches provided good guidance and feedback for those students who are willing to work on their own. With respect to coaching, one might say that it does have an effect on SAT-mathematical scores, provided that examinees also undertake independent review, which coaches may encourage, but that coaching may be of little value if no independent mathematics review is attempted. However, that most coaching would negatively affect scores seems unlikely; therefore, self-selection may be a major factor here. For example, students who both attended coaching sessions and reviewed independently may have been highly motivated. Assuming that motivation is related to test scores, these students could be expected to have received higher SAT-mathematical scores even if they had not participated in these methods of preparation, whereas less motivated students might be expected to receive lower scores. On the other hand, students who were coached, but who did not review independently, may have been less highly motivated students whose parents encouraged them to seek coaching, and who thus may have failed to achieve the potential benefits of coaching. One could conjecture a variety of other self-selection factors or situations that would be consistent with the observed data.

Interaction of Review of English with Coaching (for SAT-Mathematical Sections). Independently reviewing material from English courses also interacted significantly with coaching, showing a pattern almost identical to that of independent mathematics review with coaching. This finding probably suggests that self-selection in terms of motivation, one variable that might be expected to relate to various kinds of independent review, may be operating in both of these interactions. Students who review independently—either mathematics or English—may be more motivated than their nonreviewing counterparts. This interpretation seems more plausible than one suggesting that coaching in conjunction with reviewing English independently would produce higher SAT-mathematical scores than attending only coaching.

Interaction of Using About the SAT *with Using Test Preparation Books (for SAT-Mathematical Sections).* As described above, the effect associated with SAT-mathematical scores of choosing to complete the questions in *About the SAT* was nearly 10 points, while the corresponding estimate for using test preparation books was about -3 points. The significant interaction between these two methods shows that the positive effect of completing the questions in *About the SAT* is associated primarily with the use of test preparation books. A negligible negative effect (-2 points) was computed for using *About the SAT* when other preparation books were not used, but a +24-point effect of using the booklet was computed when other books were used as well. Alternatively, the effect of using test preparation books was +2 points if *About the SAT* were used and -23 points if it were not. Again, we could attribute these differential effects either to the combinations of methods used or to self-selection factors. Motivation, or conversely the lack of it, for the minority of students who failed to use the methods seems a likely self-selection factor.

Interaction of Review of English with Coaching (for TSWE). Overall, reviewing English or vocabulary on one's own had little effect on TSWE scores, but there was a significant negative effect (-2.8 points) associated with coaching. The significant ($p < .01$) interaction between the two methods of preparation shows that the negative effect of coaching is associated with not reviewing independently (-8.6 points) in contrast to reviewing (-0.2 points). Described in terms of the effects of reviewing independently, there was a positive effect of 9 points for those who were coached, but a negligible effect (0.6 points) for those who were not.

Alternative interpretations are again possible, depending on whether the differences result primarily from the treatments or from self-selection. On the one hand, coaching sessions may help the generally low-achieving students who attended them to focus their review on the most appropriate verbal materials for the TSWE. On the other hand, students who review independently *and* attend coaching may simply be more motivated than those who review only.

Interaction of Using Test Preparation Books with Attending In-School Review (for TWSE). Although only insignificant negative effects on TSWE scores were associated with both using test preparation books and attending in-school review sessions, these two variables interacted significantly ($p < .01$). While a positive effect of 1.5 points was computed for attending in-school review sessions for those examinees who did not use test preparation books, a negative effect (-1.9 points) was associated with those who did use such books. Alternatively, the effect of using test preparation books was 1.3 points for those who did not attend in-school review and -3.1 points for those who did. Again, differential self-selection seems more plausible here than does the effectiveness of one treatment's depending on another one, especially since negative estimates were noted here.

For those students who attended in-school review sessions, we might hypothesize the following example of one kind of self-selection that would explain the negative effect on TSWE scores of using test preparation books. These students have been selected, either by themselves or by school personnel, to attend in-school review sessions because of their deficiency in English achievement. Such students would, therefore, be expected to exhibit lower TSWE scores than unselected students, even after adjusting for their lower initial achievement.

The nature of each significant interaction effect is summarized in Table 5. It is difficult to draw any firm conclusions from these analyses with respect to which methods of preparation might differentially benefit different kinds of candidates. It does, however, seem that failing to consider the possibility of interactive effects of various methods of special preparation may be attributing to their action a simplicity that belies their apparent complexity.

Analysis II

Effects on SAT-Verbal, SAT-Mathematical, and TSWE Scores

A total of 168 students (about 20 percent of the entire group) who had used only College Board materials (either

Table 6. Patterns of Overlap of Use of Several Methods of Preparation

Method of Preparation		*Number and Percentage Using Each Other Method*								
		1	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*
Completing the questions in *About the SAT*	*N*	649	280	192	294	54	353	116	20	53
	%	100.0	43.1	30.0	45.3	8.3	54.4	17.9	3.1	8.2
Using *Taking the SAT*	*N*	280	352	97	146	28	191	53	11	27
	%	80.0	100.0	27.6	41.5	8.0	54.3	15.1	3.1	7.7
Reviewing mathematics independently	*N*	192	97	223	170	42	144	45	16	22
	%	86.1	43.5	100.0	76.2	18.8	64.6	20.2	7.2	9.9
Reviewing English or vocabulary independently	*N*	294	146	170	354	52	239	69	19	28
	%	83.1	41.2	48.2	100.0	14.7	67.5	19.5	5.4	7.9
Reviewing material from other courses independently	*N*	54	28	42	52	62	43	15	4	6
	%	87.1	45.2	67.7	83.9	100.0	69.4	24.2	6.5	9.7
Using test preparation books	*N*	353	191	144	239	43	423	91	19	39
	%	83.5	45.2	34.0	56.5	10.2	100.0	21.5	4.5	9.2
Attending review sessions in school	*N*	116	53	45	69	15	91	138	9	15
	%	84.1	38.4	32.6	50.0	10.9	65.9	100.0	6.5	10.9
Attending review or coaching sessions outside school	*N*	20	11	16	19	4	19	9	27	8
	%	74.1	30.2	59.3	70.4	14.8	70.4	33.3	100.0	29.6
Taking a previous SAT	*N*	53	27	22	28	6	39	15	8	68
	%	77.9	39.7	32.4	41.2	8.8	57.4	22.1	11.8	100.0

Table 7. Regression Equations Used to Predict SAT-V, SAT-M, and TSWE Scores for Analyses of Residuals

	Dependent Variable								
	SAT-V			*SAT-M*			*TSWE*		
Predictors	*(1)*	*(2)*	*(3)*	*(1)*	*(2)*	*(3)*	*(1)*	*(2)*	*(3)*
Intercept	11.5740	14.6970	180.0651	22.6960	33.8588	192.8934	6.5288	8.7686	8.7401
HSR	.3321	.3390	.3864	.8899	.8942	.9397	.1012	.1036	.1168
PSAT-V	8.9211	8.9140	8.9346	1.5040	1.4890	1.5088	.5659	.5621	1.7906
PSAT-M	.4394	.4838	−6.7277	7.2872	7.3178	.3824	.0964	.1116	−1.0457
Use of *About the SAT*	7.5358			6.2452			2.7232		
Use of *Taking the SAT*	−1.5565			13.9899			1.4178		
HSR squared									
PSAT-V squared									−.0135
PSAT-M squared			.0735			.0707			.0116
Multiple *R*	.8923	.8918	.8953	.8699	.8676	.8711	.7449	.7383	.7588

Note: Equations (1), (2), and (3) are based on 168 examinees who used only test sponsor-provided materials, *About the SAT* and *Taking the SAT*. The quadratic terms for HSR, PSAT-V, and PSAT-M were included when they added significantly ($p < .05$) to prediction.

Table 8. Differences Between Actual and Predicted Scores for Examinees Who Used Various Methods of Test Preparation

			Dependent Variable								
			SAT-V			*SAT-M*			*TSWE*		
Method of Preparation	*N*		*(1)*	*(2)*	*(3)*	*(1)*	*(2)*	*(3)*	*(1)*	*(2)*	*(3)*
Completing the questions	649	M	1.70	3.07	2.88	− 2.78	− 2.82	− 3.03	− 0.52	− 0.19	− 0.08
in *About the SAT*		SD	50.2	50.2	50.1	55.6	55.8	55.6	6.6	6.7	6.7
Reviewing mathematics	223	M	− 2.59	− 2.19	− 2.23	− 5.68	− 6.49	− 6.55	− 1.10	− 1.10	− 0.65
independently		SD	48.0	48.1	48.7	54.9	55.8	56.4	6.7	6.7	7.0
Reviewing English or	354	M	5.53	5.64	5.23	− 3.00	− 4.37	− 4.79	− 0.13	− 0.29	− 0.01
vocabulary independently		SD	49.9	49.8	50.3	56.5	56.9	57.1	6.7	6.7	6.8
Reviewing material from	62	M	1.53	1.98	0.00	−11.23	−11.76	−13.69	− 0.89	− 0.85	− 0.55
other courses independently		SD	43.0	43.0	47.1	57.8	59.1	59.5	7.1	7.0	7.1
Using test preparation	423	M	6.44	6.54	6.53	− 5.06	− 5.84	− 5.88	− 0.46	− 0.54	− 0.27
books		SD	52.1	52.0	51.8	54.8	55.3	55.2	7.0	7.0	7.0
Attending review sessions	138	M	− 7.02	− 6.70	− 6.67	− 7.06	− 8.72	− 8.70	− 0.84	− 0.98	− 0.94
in school		SD	49.0	49.5	50.2	55.3	55.1	54.7	6.7	6.6	6.9
Attending review coaching	27	M	1.45	1.07	− 0.30	− 7.79	− 9.69	−11.02	− 3.15	− 3.49	− 3.71
sessions outside school		SD	77.5	78.6	75.9	70.7	73.1	72.0	8.7	8.8	8.7
Taking a previous SAT	68	M	0.89	0.63	1.41	− 2.20	− 4.11	− 3.39	− 0.26	− 0.58	− 0.62
		SD	54.7	54.9	54.9	56.9	58.5	58.2	6.7	6.7	6.9
Using the booklet,	352	M	5.77	4.68	5.12	− 6.54	0.08	0.48	− 0.54	0.04	0.23
Taking the SAT		SD	49.6	49.5	49.5	54.8	55.1	55.2	6.5	6.5	6.4
No preparation	41	M	− 2.51	− 8.12	− 9.25	11.78	− 0.60	− 1.72	1.54	− 1.49	− 1.16
		SD	54.2	54.9	55.5	51.5	52.1	52.0	5.2	5.3	5.7

Note: SAT-V, SAT-M, and TSWE scores were predicted for each examinee using regression equations based on the 168 examinees who used only test-sponsor-provided materials, *About the SAT* and/or *Taking the SAT*. Column (1) is the difference between actual scores and those predicted from high school rank, PSAT-V, PSAT-M, use of *About the SAT*, and use of *Taking the SAT*. Column (2) is based on predictions from high school rank, PSAT-V, and PSAT-M. Column (3) is based on predictions from high school rank, PSAT-V, PSAT-M, and quadratic terms for each of these variables if a quadratic term contributed significantly to prediction.

About the SAT and/or *Taking the SAT*) were identified as the baseline comparison group. Of these students, 83 had used only *About the SAT*, 26 had used only *Taking the SAT*, and 59 had used both. The predictor and criterion means and standard deviations for this group were as follows: high school rank (M = 81.9, SD = 15.4), PSAT-verbal (M = 45.0, SD = 9.6), PSAT-mathematical (M = 49.7, SD = 9.6), SAT-verbal (M = 467.9, SD = 101.8), SAT-mathematical (M = 538.0, SD = 100.9), and TSWE (M = 48.1, SD = 9.4). When compared with all students in the sample, this baseline group has slightly higher status on each of the predictor and criterion variables.

Table 6 shows the overlap of use among the methods of preparation that suggested the desirability of a second, corroborative analysis. As can be seen, students who used any particular method of preparation were also very likely to prepare in one or more additional ways. Furthermore, students who used any given method were likely to differ from other students with respect to the rates at which they also used other methods. For example, whereas only 27 percent of all students reviewed mathematics on their own, 59 percent of coached students undertook such review.

Table 7 gives three separate regression equations for the baseline comparison group for each dependent variable. Table 8 contains the averages and standard deviations of the residuals for each method of preparation for each dependent variable, that is, the differences between actual scores and those predicted from each of three regression equations. As in the first analysis, the effect estimates (in this case the average of the residuals) are small. The largest average residuals for SAT-verbal scores were found for students who used test preparation books. These students exceeded predictions by about 6.5 points. Nearly all of the SAT-mathematical score average residuals were negative, regardless of the prediction equation from which expectations were established. Students who reviewed material from courses other than English and mathematics were about 12 points below their predicted SAT-mathematical scores—lower than any other group. The sole exception for SAT-mathematical scores was a positive average residual of 11.78 points for the small group of students who reportedly did not prepare at all. This positive residual was noted only for one of the three prediction equations, the one in which students' use of College Board materials was included. For the TSWE, nearly all of the residual averages were negative, with the largest (about -3.5 points) associated with attending review or coaching sessions outside of school.

Differences in the overall paterns of effect estimates may hold more information than the sizes of the individual estimates. Students who reviewed mathematics, those who reviewed other material, those who attended in-school preparation programs, and those who attended coaching sessions outside of school were farther below predictions for SAT-verbal, SAT-mathematical and TSWE scores than were the students in any other group. This situation may reflect the fact that these students had lower status than other groups on the variables used to establish predictions and possibly on important variables correlated with them. More informative, however, may be the differences between the estimates for SAT-verbal, SAT-mathematical, and TSWE scores for certain methods. For example, reviewing English has the next to largest average residuals for SAT-verbal and TSWE scores, but only the fifth largest for SAT-mathematical scores. SAT-verbal residuals were larger for those who used test preparation books than for any other group, but only sixth and fifth largest, respectively, for SAT-mathematical and TSWE scores. Perhaps the most striking difference across test sections was that the group that did not prepare at all had the largest average residuals of any group on SAT-mathematical scores, but the next to the lowest average residuals on SAT-verbal scores. This last difference strongly suggests that, in general, students' self-selection to prepare for the test is related to mathematical ability or achievement. The average PSAT-mathematical score of nonpreparing students, 50.6, which is higher than that of any group of preparing students, supports this contention.

Interaction Effects

Tables 9, 10, and 11 contain the equations for predicting, respectively, SAT-verbal, SAT-mathematical, and TSWE residuals for each method of test preparation. That is, for students who used each method, the differences between actual and predicted scores were regressed on high school rank, PSAT-verbal scores, PSAT-mathematical scores, and the students' use of each method of preparation. As the multiple R's in Tables 9-11 indicate, the residuals are predictable to a moderate degree, depending on the particular preparation method used. Multiple R's range from .21 to .57 for SAT-verbal residuals, from .16 to .70 for SAT-mathematical residuals, and from .21 to .64 for TSWE residuals.

Trait-by-Method Interactions. PSAT-verbal scores contribute significantly, and with a negative weight, to the prediction of SAT-verbal residuals for every method of preparation but one, and PSAT-mathematical scores contribute significantly for all but three methods, with positive weights. The largest weights occurred for PSAT-verbal scores in the prediction of SAT-verbal residuals for students who were coached outside of school and for PSAT-mathematical scores for students who reviewed material from courses other than English or mathematics. The direction of the weights indicated larger SAT-verbal residuals for students who scored low on the PSAT-verbal section than for those who scored high, and larger residuals for students who scored high on the PSAT-mathematical section than for those who scored low.

PSAT-verbal scores did not contribute significantly to predicting SAT-mathematical residuals. However, for four methods of preparation, PSAT-mathematical scores received significant, positive weights for predicting these

Table 9. Predictions of SAT-V Residuals for Each Method of Preparation from HSR, PSAT-V, PSAT-M, and the Use of Other Methods of Preparation

	For Students Who Used Method								
Predictor	*1*	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*
Intercept	−19.60	−26.74	− 4.17	−15.16	14.30	−32.29	6.14	110.28	−80.37
HSR	.1317	.1219	− .4533	.0027	− .8188	.3111	− .0121	1.3042	1.0219
PSAT-V	− .7714**	− .7649*	− 1.2687**	− 1.3261**	− 1.8420*	− 1.3062**	− 1.6824**	− 6.1778*	− 1.7659
PSAT-M	.8407**	1.0413**	1.8518**	1.5899**	3.0411**	1.4608**	.9186	.8877	1.2556
1. Completing the questions in *About the SAT*	−	− .86	6.35	4.32	1.74	.94	22.66	− 18.77	8.86
2. Using the booklet, *Taking the SAT*	3.00	−	5.95	4.14	− 8.08	6.53	− .49	2.16	4.15
3. Reviewing mathematics independently	− 7.77	− 4.71	−	−10.59	−12.64	− 2.75	−24.19*	2.13	− 7.47
4. Reviewing English or vocabulary independently	6.78	4.79	− 7.42	−	−20.35	1.21	8.91	60.95	− 1.87
5. Reviewing material from other courses independently	− 4.51	− 8.05	− 4.93	− 5.06	−	1.27	− 8.15	20.32	−25.13
6. Using test preparation books	8.50*	14.06*	16.90*	6.31	30.30*	−	6.00	− 27.69	18.15
7. Attending review or coaching sessions outside school	− 7.30	−12.52	−26.64**	−12.68	−28.94*	−10.26	−	−30.23	−13.32
8. Attending review or coaching sessions outside school	− 5.89	1.92	− 1.60	10.10	−16.82	− 6.26	− 2.03	−	34.73
9. Taking a previous SAT	− .12	− 1.08	− 3.04	− 4.41	− 9.87	.75	−11.75	27.52	−
Multiple *R*	.21**	.25*	.39**	.32**	.57**	.29**	.38*	.65	.43

Note: Residuals are differences between actual scores and those predicted from HSR, PSAT-V, PSAT-M, and PSAT-M squared (which added significantly to prediction) for 168 examinees who used only test-sponsor-provided materials.

*$p < .05$

**$p < .01$

Table 10. Predictions of SAT-M Residuals for Each Method of Preparation from HSR, PSAT-V, PSAT-M, and the Use of Other Methods of Preparation

	For Students Who Used Method								
Predictor	*1*	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*
Intercept	−23.95	−16.82	−25.14	−40.75	−83.00	−52.31	− 7.40	78.68	−68.60
HSR	− .2479	− .1005	− .6600*	− .4923*	− .1143	− .3532	− .4463	.2398	− .4164
PSAT-V	− .2079	− .1776	− .4664	− .4462	.5082	− .0548	.1377	− 4.3184	.0102
PSAT-M	.9111**	.4563	1.7372**	1.5992**	1.9056	1.0777**	.4194	1.2065	1.5278
1. Completing the questions in *About the SAT*	–	14.62	19.31	20.78**	−21.14	25.04**	24.92	12.46	17.05
2. Using the booklet, *Taking the SAT*	10.39*	–	17.78*	7.04	23.22	10.08	− 2.41	17.81	25.08
3. Reviewing mathematics independently	.40	6.04	–	− 2.73	−13.57	− 4.56	−16.72	56.03	− 1.67
4. Reviewing English or vocabulary independently	3.68	− 2.29	− 6.86	–	− 7.64	5.20	.89	32.99	17.49
5. Reviewing material from other courses independently	−15.08	.52	−19.26*	−14.67	–	−12.78	4.29	−20.46	−23.11
6. Using test preparation books	1.88	− .80	− 6.90	.05	−11.87	–	− 8.26	−26.15	9.08
7. Attending review sessions in school	− 2.39	−13.23	−18.18*	− 8.51	1.82	− 6.48	–	−51.34	−13.56
8. Attending review or coaching sessions outside school	.06	− 2.84	25.55	17.33	−13.06	1.68	−21.48	–	−39.16
9. Taking a previous SAT	3.59	13.76	8.49	14.50	13.64	10.59	−11.25	−44.30	–
Multiple *R*	.18*	.16	.38**	.29**	.52	.27**	.27	.70	.44

Note: Residuals are differences between actual scores and those predicted from HSR, PSAT-V, PSAT-M, and PSAT-M squared (which added significantly to prediction) for 168 examinees who used only test-sponsor-provided materials.

*$p < .05$

**$p < .01$

Table 11. Predictions of TSWE Residuals for Each Method of Preparation from HSR, PSAT-V, PSAT-M, and the Use of Other Methods of Preparation

	For Students Who Used Method								
Predictor	*1*	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*
Intercept	.80	.21	−1.73	1.47	4.10	−2.22	5.40	1.89	3.47
HSR	− .0551**	− .0489	− .0735*	− .0805**	− .1304	− .0711**	− .1144**	.0753	− .0071
PSAT-V	− .0606	− .0200	− .0543	− .0529	− .0390	− .0530	− .0861	− .4801	− .3204*
PSAT-M	.1227**	.0756	.1762**	.1436**	.1516	.1798**	.1605*	.0860	.2379
1. Completing the questions in *About the SAT*	–	1.74*	1.84	.41	1.12	.87	1.14	1.52	2.14
2. Using the booklet, *Taking the SAT*	1.09*	–	.72	.73	.40	1.13	− .04	−1.19	−1.74
3. Reviewing mathematics independently	− .27	− .63	–	−1.07	1.08	− .32	−2.55*	3.66	1.94
4. Reviewing English or vocabulary independently	.47	.60	−1.09	–	−3.23	1.13	1.57	7.52	− .79
5. Reviewing material from other courses independently	− .49	− .12	.50	− .72	–	.46	−2.08	6.32	−2.35
6. Using test preparation books	− .16	.42	.72	.78	2.82	–	−2.64*	−3.33	− .92
7. Attending review sessions in school	− .95	−2.17*	−3.00*	− .00	−4.46	−1.89*	–	−2.19	−3.28
8. Attending review or coaching sessions outside school	−2.94	−2.31	− .10	− .48	.74	−2.98	−3.14	–	−6.07*
9. Taking a previous SAT	.61	−1.78	.12	− .87	−1.53	.14	−1.13	−2.59	–
Multiple *R*	.21**	.23	.27	.24*	.39	.28**	.42**	.64	.47

Note: Residuals are differences between actual scores and those predicted from HSR, PSAT-V, PSAT-M, and PSAT-M squared (which added significantly to prediction) for 168 examinees who used only test-sponsor-provided materials.

*$p < .05$

**$p < .01$

residuals. Each of the other weights for PSAT-mathematical scores was also positive, although not significant. High school rank received negative weights in predicting SAT-mathematical residuals for all but one method, and it received significant negative weights for students who used each of two independent review strategies: reviewing mathematics and reviewing English or vocabulary.

PSAT-mathematical scores also received positive weights (five of which were significant) for predicting TSWE residuals, and high school rank received significantly negative weights for the same five methods.

Method-by-Method Interactions. Several interactions were signaled by the significant contributions made to the prediction of residuals by several variables indicating whether or not students used other methods of preparation. For SAT-verbal scores, the effects of attending in-school preparation programs were significantly greater for students who did not review mathematics independently than for those who did. The effects of using *About the SAT*, of using *Taking the SAT*, of reviewing mathematics independently, and of reviewing independently material from other courses were greater for students who also used test preparation books than for those who did not. Finally, the effects of reviewing mathematics and reviewing material from other courses were significantly greater if students had not attended an in-school preparation program than if they had.

For SAT-mathematical scores, the effects of reviewing English or vocabulary independently and of using test preparation books were greater ($p < .01$) for students who also used *About the SAT* than for those who did not. The effects of using *About the SAT* and of reviewing mathematics independently were greater when students also used *Taking the SAT*. The effects of independent mathematics review were less, however, when students also either reviewed material from other nonmathematics, non-English courses or attended review sessions in school.

For the TSWE, the effects of using *About the SAT* were greater when students also used *Taking the SAT*, and conversely, the effects of *Taking the SAT* were greater when *About the SAT* was used. The effects of *Taking the SAT*, of reviewing mathematics, and of using test preparation books were less if students had also attended in-school review sessions [and the effects of in-school review were significantly less ($p < .05$)] if students had also reviewed mathematics or used test preparation books. Finally, the effects of taking the SAT previously were significantly less ($p < .05$) for students who had attended coaching sessions outside of school than for those who did not.

Comparison of Findings with Results of Previous Studies

A recent study conducted by the Federal Trade Commission (FTC) (1979) found that the effects of commercial coaching were related to the particular coaching school attended. Whereas the effects associated with one school were calculated to be about a 30-point increase on SAT-verbal scores and about a 20-point increase on SAT-mathematical scores, the effects associated with attending another commercial coaching school were negligible. The analyses presented here also suggest that the effects associated with coaching obtained outside of school, but not necessarily commercial coaching, were negligible. However, the students in the present study seem to have been selected on a different basis from that of students who attended the schools investigated by the FTC. The coached students in the FTC study had significantly higher class ranks, better grades in mathematics and English courses, and greater parental incomes than uncoached students. Candidates who were coached also reported that they expected to take more mathematics courses than did uncoached students. On the other hand, in this study, coached students reported significantly *fewer* years of mathematics and lower grades in mathematics courses than uncoached students.

In another study, Alderman and Powers (1980) investigated the effects of programs offered by secondary schools. Using random assignment to treatments, so that the effects of selection were not confounded with the effects of test preparation, the authors estimated an average effect of 8 points for 8 high school preparation programs designed to improve SAT-verbal scores. This estimate, which, because of random assignment, is not confounded with self-selection, seems more realistic than the present study's significantly negative effect estimate on SAT-verbal scores for in-school review programs. It is likely that the negative estimate obtained in the current study is inaccurately low, reflecting the failure to adjust adequately for the in-school preparation group's low achievement/ability status.

SUMMARY AND DISCUSSION

A variety of test preparation resources and methods are available to, and used by, SAT candidates. In the present study, the effects on SAT and TSWE scores associated with the selection and use of each of several methods of preparation were small—less than 10 points on the 200-800 SAT score scale and less than 2 points on the 20-60 TSWE scale. With the exception of the estimate of the effect of receiving the booklet, *Taking the SAT*, which was distributed randomly to half the sample, all of the estimates reflect both use and selection.

Generally, in the sample investigated here the groups of examinees who reported using each method of preparation had somewhat lower average PSAT scores and, in many cases, lower average ranks in their high school classes than did nonpreparing examinees. An analysis of background characteristics suggested that electing to use each of the methods of preparation is also related to other factors, such as concern or anxiety about test performance, previous

below-average performance in mathematics for those who reported being coached outside of school, and minority status for those who reviewed independently. Examinees who chose certain methods exhibited, by their reports of nervousness, more concern than nonpreparers about their test performance. Those who were coached outside of school had taken fewer mathematics courses and had acnieved lower grades in the mathematics courses that they had taken. Coached students also reported higher levels of education for their mothers. Candidates who reported an independent review of either mathematics or English were more likely to be from minority groups.

The effects of selection were also suggested by the fairly strong relationships between the size of the effect estimates, for both SAT-mathematical and TSWE scores, and each group's relative standing on the three covariates–PSAT-verbal scores, PSAT-mathematical scores, and HSR. The rank order correlations between the size of the effect on SAT-mathematical scores determined in one analysis for each method of preparation and the covariate means of the groups that used each method were large (.87, .72, and .73) and statistically significant ($p < .05$). That is, the higher the group's PSAT-verbal scores, PSAT-mathematical scores, or HSR, the higher was the estimate of the particular method of preparation. The same was true for the relationship between TSWE effect size and each group's relative standing on each of the three covariates (.75, .86, and .75). These strong relationships between effect size and relative standing on covariates suggest that, despite the very high multiple correlations, adjusting for only PSAT-verbal scores, PSAT-mathematical scores, and HSR may not have been sufficient, and that individual effect estimates are, in general inaccurately high (or low) depending on the initial status of the preparers. Lord (1967, 1969) and others have written extensively about the problems involved in attempting to adjust for such preexisting group differences in the analysis of covariance. Furthermore, because the bases for selection appeared to differ from method to method, some estimates probably reflect more selection bias than others. That some estimates are inaccurately low is supported by the fact that they are negative, a finding that carries the seemingly implausible implication that some kinds of preparation could actually be harmful. In short, the evidence is strong here for the existence of other unconsidered factors that are related both to choosing to prepare for the SAT and to test scores.

Significant interactions between some variables suggested that the effects of using various methods of preparation are not necessarily uniform across levels of ability or achievement, and that their effects may differ depending on whether they are used in conjunction with other methods. Alternatively, the effects of, and probably the bases for, self-selection may not be uniform across ability/achievement levels or may depend on examinees' selection of other methods of preparation.

The present study, like other studies of the effects of test preparation, has limitations, perhaps the greatest of which is its reliance on observational, instead of experimental, data. Several instances of possibly biased estimates of effects resulting from the analysis of these data have been discussed above. Perhaps the most significant results of the present study stem from its limitations, which may have implications for the design of future studies of the effects of test preparation. In view of the numerous alternative methods of preparation and, apparently, the considerable entanglement of the effects of self-selection and use, the following strategies seem appropriate for further studies.

Random Assignment

Random assignment to test preparation methods seems especially desirable for studies of the effects of special preparation, in which effects may be quite small relative to the potential effects of self-selection factors. However, given the number of methods and resources available, random assignment is undoubtedly difficult to implement, and the resulting control conditions are equally difficult to maintain. Students who are denied access to a specific method of preparation, even temporarily, may choose to substitute any of a number of alternatives. Thus, in any randomized study, checks probably should be made on examinees' use of test preparation methods other than those specifically under investigation. In addition, given the possibility of joint effects of some methods of preparation, simultaneous random assignment to more than one method (or to more than one level) of preparation would provide better evidence regarding the interactive effects of various methods of preparation.

Description of Preparation Methods

The methods of preparation, or treatments, need to be thoroughly described. The simple categorization used in the present study could be improved because of the following.

1. There is considerable overlap among methods. For example, commercially available test preparation books or familiarization materials provided by the test sponsor may be the primary instructional resources in a coaching program or in an in-school review program. Self-study or independent review may also be an important component of such programs.

2. Some preparation methods, although nominally the same, may focus on the different sections of the test, or may have varying objectives (e.g., anxiety reduction or test familiarization) and therefore may need to be evaluated on different bases.

3. Considerable variation may exist in the scope and quality of resources within a given category of preparation. For example, there are numerous commercially available test preparation books and coaching courses that vary in both the quantity and the quality of material imparted and

in the amount of in-class instruction or self-study required. Other methods, such as candidates' independent review, surely exhibit similar variability with respect to both quantity and quality.

4. The effects of the selection and use of various preparation resources may depend on the order in which they are used. For example, taking the SAT for practice may help candidates to focus on the most relevant features of some other method of preparation, or attending a coaching or review session may help to guide students' independent review.

5. Although factual information may be reported with considerable accuracy (Baird 1976), some verification of students' self-reports would add greater credibility to future studies. This seems important in light of the potential for candidates' misinterpreting survey questions because of the hard-to-eliminate ambiguity present in questions about test preparation methods.

Assessing Factors in Self-Selection

In the absence of random assignment to treatments, reliably measuring the factors that are related to test scores and to the selection of preparation methods is critical for obtaining less biased estimates of effects. The measurement of possible selection factors that may be positively related to test scores (e.g., motivation and career aspirations) and the measurement of those that may be negatively related to test scores (e.g., extreme test anxiety and low previous achievement) should be given equal consideration.

In addition, selection factors might be also assessed even more directly by asking examinees about the reasons for choosing particular methods of preparation. Although the assumption is usually made that students *self*-select test preparation treatments, other kinds of selection may also operate for some preparation methods. Some students may be selected by secondary schools or parents to attend special preparation, and may in fact be reluctant participants rather than self-selected ones.

Nontest Score Effects

The effect on total test scores has been the major concern in most studies of total test preparation, and some studies have addressed the effects of test preparation on specific test item types. Other nontest score effects that test preparation might produce are (1) reduced anxiety and less apprehension about the test and (2) changes in the reliability and/or validity of test scores resulting from reduction in error variance. These effects may be equally important and may deserve study.

Sample Selection

Because effects may differ according to the characteristics of the study sample, future investigators may wish to consider the effects of preparation for diverse groups of students with respect to such factors as motivation to perform well on the test, previous achievement and ability, and prior test-taking history. The present study includes only students who had previously taken the PSAT, a highly relevant preparation for the SAT. In addition, all of the examinees in the present sample were juniors, who might have had different views from those of high school seniors concerning the SAT and preparing for it.

Concluding Remarks

In pointing out the difficulties involved in determining the effects of special preparation on SAT scores, the present study emphasizes the complexity of the special preparation issue and, perhaps, underscores the futility of attempting a single "block-buster" study of special preparation. The study has also generated some hypotheses that might be investigated in a more rigorous fashion than was possible here. For example, the effect of each individual method of preparation noted here was small, and it was usually overshadowed by the effects of self-selection factors, even when strong covariates were used to control for preexisting differences between groups. Nevertheless, some combinations of preparation methods were associated with practically and statistically significant score effects. Because of the possible confounding of these estimates with self-selection factors, further consideration of some combinations seems warranted. One case in point is the interaction of coaching outside of school with the independent review of mathematics, a combination that appeared to have a relatively large effect on SAT-mathematical scores.

The present study also suggests that independent study —of either English or mathematics, or of test preparation books and test familiarization materials—may have an impact on test scores that is possibly as great as the effect of less widely available modes of preparation. To provide some balance to the considerable attention that has been afforded coaching programs recently, further and more rigorous study of the degree to which SAT scores may be enhanced by independent study is recommended, since the extent to which such study may produce score increases that are comparable to those obtained by the (usually) more expensive coaching programs bears directly on the question of equity of access. The differences between the effects of coaching programs and the corresponding effects of independent study may be more germane than the effects of coaching per se.

Finally, the present study suggests, as have other previous studies, that motivation may be a potentially important self-selection factor in quasi-experimental studies of special preparation. Here, the apparent joint effect of particular methods of preparation was often greater than the sum of the effects of each individual method, suggesting perhaps that particular combinations are especially bene-

ficial. However, given the nature of the methods that interacted in this manner, a more plausible alternative explanation is that examinees who use each of two methods may be more highly motivated than examinees who use only one of the methods. The importance of motivation as a self-selection factor and its relationship to SAT scores therefore seem to deserve further attention. Conversely, that the joint effects of some methods of preparation were less than the sum of effects of individual methods suggests that using particular combinations of methods may not be very efficient, possibly because of redundancies in the emphases and effects of the methods.

REFERENCES

Alderman, D. L., and Powers, D. E. 1980. The Effects of Special Preparation on SAT-Verbal Scores. *American Educational Research Journal* 17: 239–251.

Baird, L. L. 1976. *Using Self-Reports to Predict Student Performance.* Research Monograph No. 7. New York: College Entrance Examination Board.

Campbell, D. T., and Boruch, R. F. 1975. Making the Case for Randomized Asssignment to Treatment by Considering the Alternatives: Six Ways in Which Quasi-Experimental Evaluations in Compensatory Education Tend to Underestimate Effects. In *Evaluation of Experience: Some Critical Issues in Assessing Social Programs,* A. Lumsdaine and C. A. Bennett, eds. New York: Academic Press.

Campbell, D. T., and Erlebacher, A. 1970. How Regression Artifacts in Quasi-Experimental Evaluations Can Mistakenly Make Compensatory Education Look Harmful. In *Disadvantaged Child,* Vol. 3, J. Hellmuth, ed. New York: Brunner/Mazel.

Coffman, W. E., and Parry, M. E. 1967. Effects of an Accelerated Reading Course on SAT-V Scores. *Personnel and Guidance Journal* 46: 292–296.

Cronbach, L. J., and Snow, R. E. 1977. *Aptitudes and Instructional Methods.* New York: Irvington Publishers.

Dyer, H. S. 1953. Does Coaching Help? *College Board Review* 19: 331–335. (Reported in French, J. W., and Dear, R. E. 1959. Effect of Coaching on an Aptitude Test. *Educational and Psychological Measurement* 19: 319–330).

Evans, F. R., and Pike, L. W. 1973. The Effects of Instruction for Three Mathematics Item Formats. *Journal of Educational Measurement* 10: 257–272.

Federal Trade Commission. 1979. *Effects of Coaching on Standardized Admission Examinations: Revised Statistical Analyses of Data Gathered by Boston Regional Office of Federal Trade Commission.* Washington, D.C.: Federal Trade Commission, Bureau of Consumer Protection.

Frankel, E. 1960. Effects of Growth, Practice, and Coaching on Scholastic Aptitude Test Scores. *Personnel and Guidance Journal* 38: 713–719.

French, J. W. 1955. *The Coachability of the SAT in Public Schools* (RB 55-26). Princeton, N.J.: Educational Testing Service.

French, J. W., and Dear, R. E. 1959. Effect of Coaching on an Aptitude Test. *Educational and Psychological Measurement* 19: 319–330.

Lord, F. M. 1967. A Paradox in the Interpretation of Group Comparisons. *Psychological Bulletin* 68: 304–305.

Lord, F. M. 1969. Statistical Adjustments when Comparing Pre-Existing Groups. *Psychological Bulletin* 72: 336–337.

Marron, J. E. 1965. *Preparatory School Test Preparation: Special Test Preparation, Its Effect on College Board Scores and the Relationship of Affected Scores to Subsequent College Performance.* West Point, N.Y.: Research Division, Office of the Director of Admissions and Registrar, United States Military Academy.

Messick, S. 1980. *The Effectiveness of Coaching for the SAT: Review and Reanalysis of Research from the Fifties to the FTC.* Princeton, N.J.: Educational Testing Service.

Pallone, N. J. 1961. Effects of Short-Term and Long-Term Developmental Reading Courses upon SAT Verbal Scores. *Personnel and Guidance Journal* 39: 654–657.

Powers, D. E. 1981. Students' Use of and Reactions to Alternative Methods of Preparing for the SAT. *Measurement and Evaluation in Guidance* 14: 118–126.

Powers, D. E., and Alderman, D. L. 1979. *The Use, Acceptance, and Impact of Taking the SAT–A Test Familiarization Booklet* (RDR 78-79, No. 6, and ETS RR-79-3). Princeton, N.J.: Educational Testing Service.

Roberts, S. O., and Oppenheim, D. B. 1966. *The Effect of Special Instruction upon Test Performance of High School Students in Tennessee* (CB RDR 66-7, No. 1 and ETS RB 66-36). Princeton, N.J.: Educational Testing Service.

Stroud, T. W. F. 1980. *Reanalysis of the Federal Trade Commission Study of Commercial Coaching for the SAT* (RR 80-10). Princeton, N.J.: Educational Testing Service.

Whitla, D. K. 1962. Effect of Tutoring on Scholastic Aptitude Test Scores. *Personnel and Guidance Journal* 41: 32–37.